

**Study report to the
Directorate General for Translation
of the European Commission
Final version**

DGT-ML-STUDIES 08

Study on the size of the language industry in the EU

17th August 2009

The Language Technology Centre Ltd

5-7 Kingston Hill

Kingston upon Thames

Surrey KT2 7PW

United Kingdom

www.langtech.co.uk

Executive Summary

The present document is the final report of a six-month study conducted by LTC which addresses the need of a comprehensive and systematic analysis of the size of the language industry in the European Union. As well as presenting a snapshot of the current situation, estimates are provided of how the industry will develop in the future.

Thorough secondary and primary research allowed to propose an estimated value of the language industry within the European Member States of **8.4 billion €** for 2008. This figure comprises the industry sectors of translation, interpreting, software localisation and website globalisation, language technology tool development, language teaching, consultancy in linguistic issues and organisation of international conferences with multilingual requirements. In addition, it includes language-related activities performed in corporate environments.

Market forecast

Taking into consideration all sources of information as explained in more detail below, the **annual compounded growth rate** was estimated at **10%** minimum over the next few years, resulting in an approximate value of the language industry of **16.5 billion €** in 2015. The language industry seems to be less affected by the financial crisis than other industry sectors. Where turnovers from multilingual business activities have been negatively impacted, this has been mainly in the case of individuals and micro-companies dependant on a small number of clients, a quick recovery and continued steady growth of the market is forecasted. Although this forecast is highly speculative due to fragmented information available on the industry, it should be considered conservative. Further research is required to confirm that the real value of the language industry can be expected to be well above 20 billion € by 2015.

In the future, the impact of multilingual competence on economic productivity should be analysed systematically in order to arrive at a percentage of the value generated by the language industry on the economic output of European countries and Europe as a whole. Currently, multilingualism and its importance is underestimated. When implementing subsidiaries in other countries, staffing, office location and facilities are of primary concern to the management, whereas proper presentation of a foreign company's products and services in the target country's language and the impact of this aspect whilst competing with national competitors is often underestimated, especially by SMEs with limited funds. Failure abroad is usually attributed to general management weakness rather than target language related failures. In many cases, it is incorrectly assumed that localising a website into a language of a target country is sufficient to generate sales.

The increasingly better understanding of the importance of the language industry for successful globalisation is reflected in the fact that investment companies are showing increasing interest and are buying stakes in language companies.

It is nevertheless necessary to actively and visibly address and promote the importance of overcoming language barriers in order to ensure economic success for European companies within and outside Europe much more than is currently the case.

Results and Conclusions

Data collected through the sources mentioned in the study methodology section of this summary was analysed by sector and country in order to draw significant conclusions. All materials consulted were then classified and inserted into a searchable knowledge base delivered to the European Commission with the final report. In addition, for every Member State a comprehensive country fact sheet was created, containing statistical information about the Member State, a list of professional organisations contacted, information about the national statistics office, a list of authorities contacted and whether they have provided any data. Most importantly, the fact sheets contain a list of the publications collected, divided by sector and sub-sector. In some cases, the main findings of the publications are summarised in the single fact sheets.

For the majority of countries, the data retrieved allowed for an estimate of the value of translation and interpreting in the specific country. Only in very few cases a country-specific estimate of the value of other sectors was feasible as a result of official sources of information.

In the case of translation and interpreting, the figures provided for individual countries were combined and the missing values (for those countries where no data was available) were estimated. This led to an assumed value of the translation and interpreting sector, comprising software localisation and website globalisation activities, of **5.7 billion €** in 2008.

For the remaining sectors, relevant publications were taken into account to estimate the total value. For the year 2008, the sector of language technology tools was estimated at **568 million €**, the sector subtitling and dubbing at **633 million €**, language teaching at **1.6 billion €** and multilingual support within conference organisation at **143 million €**. It is important to point out that all figures above include the value generated in corporate environments by employees directly or indirectly responsible for multilingual data processing.

For those sectors where information is available in some Member States, it is believed that the same kind of data could be produced in the same format by the other Member States as well, leading to a solid base of comparison for more reliable future market estimates.

Due to the lack of accurate data as explained later in this summary, the figures derived about the size and the volume of the language industry in Europe are often based on assumptions and must therefore be considered highly speculative.

Some of the additional findings that emerged from secondary and primary research are summarised as follows:

- As regards the sector of translation and interpreting, market consolidation continues and this trend is likely to remain constant in the future. As a result, the currently fragmented nature of the language industry will continue to consolidate into larger commercial entities. Data shows that a smaller number of increasingly larger players seem to dominate the field with growing presence in Eastern Europe.
- Entry barriers to the field of translation and interpreting are low. The main consequence is increasingly fierce and sometimes unfair competition, as well as a decrease in prices combined with a decline in quality levels. The new EN15038 certification designed to counteract this trend seems to be well implemented across Europe but only addresses quality issues partially and therefore appears to require amendments.
- In order to deliver appropriate output as required for successful foreign market penetration achieved by the clients of multilingual experts and LSPs, and to prevent unqualified players from entering the language industry, the results confirm that the profile and image of language experts need to be raised and their remuneration increased accordingly. Growing needs for specialised domain and technological skills should be better addressed by universities and colleges offering linguistic qualifications.
- In the sector of language technology tool development, the focus continues to be on linguistic tools. Linguists' resistance to machine translation is decreasing. It seems very likely that the use of machine translation will grow to cater for exponentially rising translation needs in increasingly globalised contexts, combined with a considerable lack of properly skilled human resources. Visibility and acceptance of machine translation is supported by the efforts and availability of tools free of charge by Google and Microsoft. The number one in machine translation continues to be Systran, followed by Google in second place. Machine translation capability is likely to be increasingly embedded in multilingual workflows. Translation support tools and workflow automation are likely to be used more widely, supporting and increasing the speed and the quality of multilingual output, at the same time decreasing the cost of human efforts. Tools currently developed and designed appear to become more and more sophisticated as professional tools for highly specialised language experts. It is worth mentioning that the majority of tools are developed in Europe and that this strength should be supported further.
- The sector of subtitling is in clear need of regulations on a European level in order to counteract trends such as peer-to-peer subtitling and outsourcing to Asian countries, which result in decreasing quality levels.
- English, German, Spanish, French and Italian continue to be the most widely used languages throughout Europe. The use of regional languages is supported and

increasing, e.g. in countries like the UK (Welsh and Scottish Gaelic), Ireland (Irish Gaelic has become an official language in the EU), Spain and Malta. Crowd sourcing is likely to contribute to the survival and strengthening of regional and minority languages.

In addition to estimating the turnovers for the single sectors, some countries were highlighted as “best practice examples” on the basis of the reliability and accuracy of data provided.

- For the sector translation and interpreting: Belgium, Denmark, Estonia, Finland, Germany, Italy, Lithuania, Romania and Slovenia.
- For the sector of subtitling and dubbing: Slovakia.
- For private language tuition: the Czech Republic, Finland, Slovakia.
- For institutional language tuition: Austria, Slovakia, the United Kingdom.
- For language technology tool development: France.

For the remaining sectors, no countries could be pointed out as examples of best practice as regards data provision.

Recommendations

Given the issues encountered at data provision level with the national authorities, it is considered of utmost importance to ensure systematic and consistent reporting on language matters across Europe by introducing accurate measures at national statistics offices in all Member States. In countries with a system of autonomous regions, such as Germany, national governments should be encouraged to ensure consistent reporting across their regional authorities in order to present unified statistics for the entire country. Once comparable statistics are generated on a European level, they can provide a basis for further evaluation and foreign language policy planning.

The primary research was based on more than 1000 valid responses across Europe. Whilst this leads to reliable analysis results at European level, the respondent rate per country (on average 40 and a lot less for the smallest countries) across the 27 Member States should be increased in future studies across all sectors to allow for more precise conclusions at national level or per sector in a specific country.

According to industry experts, in 2008 Europe had a major share of the global revenues generated by multilingual services. This exceptionally high share of the market underlines the potential and the rich source of income represented by the language industry to the EU economy. Research shows that the language industry has the highest growth rate of all European industries in Europe. In order to be able to capitalise on this asset, in an

environment of increasing globalisation and growing need for multilingualism, the language industry needs to be better understood, sustained and stimulated.

In order to be exploited to the maximum, a report of this kind should reflect the dynamic character of the language industry and hence be regularly updated to represent a solid basis of comparison to study the developments of the industry, draw appropriate conclusions and take measures in terms of political planning and commercial investment. Without doubt, additional valuable information can be provided by knowledgeable readers of this report. It is therefore recommended to introduce a certain element of crowd-sourcing and joint efforts by interested parties, taking advantage of the latest web technology available via the internet, allowing suitable individuals and organisations to regularly contribute to and take advantage of updated information. This is of course subject to regular monitoring and validation by experts within the European Commission or by suitable subcontractors.

Study methodology

In order to arrive at an estimate on the total size of the language industry in the European Union, systematic and comprehensive searches were performed across a variety of sources of information such as statistics databases from the European Union, publications by professional associations, databases and other published market research reports. For this purpose, more than 200 national, European and international professional associations were contacted with the request to provide published studies or statistics about the language industry. In addition, in the course of the study more than 100 national authorities were contacted with the request to supply data available on the language industry (or specific sectors) in their respective country. The types of authorities that were contacted include Statistics Offices, Business Registers, Ministries of Education, Tax offices, Ministries of Finance and other entities.

In addition to performing secondary research, a comprehensive questionnaire was developed and distributed among several thousand individuals, language service providers and language service departments across the European Union. In total, more than 1000 participants responded to the questionnaire.

Limitations of the study

Data publicly available at Eurostat and national statistics offices proved to be difficult to obtain and only available for some of the countries and sectors. Inconsistencies at national level propagated to a regional level in the case of the language teaching sector in Germany, where responsibilities lie with regional governments in the *Länder*. Data was therefore provided in a variety of formats and hence not comparable.

As regards data made available by national authorities, statistics proved difficult to be compared since classification codes differed among countries, as did the levels of detail of the

data supplied. In addition, response times were sometimes quite long, which significantly reduced the time available for data analysis.

Notwithstanding the above obstacles, thorough primary and secondary research and considerable work creating a well structured and detailed report have led to results that are more substantial and better founded than other reports that are usually focussed on only one or some of the highly complex aspects of the language industry. Moreover, this report can be used as a starting point for future work according to the results, conclusions and recommendations presented above